

Improving the Correlation Lower Bound for Simultaneous Orthogonal Matching Pursuit

Jean-François Determe^{*†} Jérôme Louveaux[†] Laurent Jacques[†] François Horlin^{*}

October 14, 2016

Abstract

The simultaneous orthogonal matching pursuit (SOMP) algorithm aims to find the joint support of a set of sparse signals acquired under a multiple measurement vector model. Critically, the analysis of SOMP depends on the maximal inner product of any atom of a suitable dictionary and the current signal residual, which is formed by the subtraction of previously selected atoms. This inner product, or correlation, is a key metric to determine the best atom to pick at each iteration. This paper provides, for each iteration of SOMP, a novel lower bound of the aforementioned metric for the atoms belonging to the correct and common joint support of the multiple signals. Although the bound is obtained for the noiseless case, its main purpose is to intervene in noisy analyses of SOMP. Finally, it is shown for specific signal patterns that the proposed bound outperforms state-of-the-art results for SOMP, and orthogonal matching pursuit (OMP) as a special case.

1 Introduction

The recovery of signals possessing a sparse representation in some orthonormal basis Ψ , *i.e.*, signals fully expressed using a limited number of vectors from Ψ , acquired by means of a linear measurement process is a problem that has gained in popularity in the last decade with the emergence of the *compressive sensing* (CS) [5, 14] theory. This paper analyzes simultaneous orthogonal matching pursuit (SOMP) [22] for this sparse signal recovery problem involving possibly more than one sparse signal to be retrieved.

1.1 Signal model

Let us now define our models of interest. For the sake of clarity, we assume below that $\Psi = I$ but all our results can easily be adapted to the general case. Using $[n] := \{1, \dots, n\}$, we define the support of any vector \mathbf{x} as $\text{supp}(\mathbf{x}) = \{j \in [n] : x_j \neq 0\}$ with \mathbf{x} being s -sparse whenever $\|\mathbf{x}\|_0 := |\text{supp}(\mathbf{x})| \leq s$. In this context, x_j is the j th entry of \mathbf{x} while $|\cdot|$ denotes the cardinality. In a single measurement vector (SMV) signal model [15], we consider a $|\mathcal{S}|$ -sparse signal \mathbf{x} whose support is \mathcal{S} and the corresponding measurement vector $\mathbf{y} \in \mathbb{R}^m$ gathering measurements of \mathbf{x} :

$$\mathbf{y} = \Phi \mathbf{x}, \quad (1)$$

^{*}Jean-François Determe and François Horlin are with the OPERA Wireless Communications Group, Université Libre de Bruxelles, 1050 Brussels, Belgium. E-mail: jdeterme@ulb.ac.be, fhorlin@ulb.ac.be. Jean-François Determe is funded by the Belgian National Science Foundation (F.R.S.-FNRS).

[†]Laurent Jacques, Jérôme Louveaux, and Jean-François Determe are with the ICTTEAM département, Université catholique de Louvain. E-mail: laurent.jacques@uclouvain.be, jerome.louveaux@uclouvain.be. Laurent Jacques is funded by the Belgian National Science Foundation (F.R.S.-FNRS).

where $\Phi \in \mathbb{R}^{m \times n}$ describes the linear measurement process being used. We find convenient to refer to the columns of the measurement matrix Φ as *atoms*. This terminology is usually employed when dealing with dictionaries, which implicitly exist in our signal model. For $\mathcal{S} := \text{supp}(\mathbf{x})$, Equation (1) indeed rewrites $\mathbf{y} = \sum_{j \in \mathcal{S}} x_j \phi_j$ where ϕ_j denotes the j th column (or atom) of Φ . Thus, recovering \mathcal{S} is equivalent to determining which set of $|\mathcal{S}|$ columns from Φ enables one to fully express \mathbf{y} using the proper linear combination.

Even for $m < n$, it can be shown that several algorithms of reasonable complexity can recover any sufficiently sparse signal \mathbf{x} provided that the matrix Φ satisfies some properties. Among them, the restricted isometry property (RIP) [5] is probably one of the most ubiquitous in the CS literature. A matrix Φ satisfies the RIP of order s with restricted isometry constant (RIC) δ_s if and only if $\delta_s \in [0, 1)$ is the smallest δ such that

$$(1 - \delta)\|\mathbf{u}\|_2^2 \leq \|\Phi\mathbf{u}\|_2^2 \leq (1 + \delta)\|\mathbf{u}\|_2^2 \quad (2)$$

is true for all s -sparse vectors \mathbf{u} . For a given measurement matrix Φ , the RIC determines how close the ℓ_2 -norms of any s -sparse signal and its associated measurement vector are.

The signal sensing model (1) can be generalized by considering the associated multiple measurement vector (MMV) signal model [15]

$$\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_K) = \Phi (\mathbf{x}_1, \dots, \mathbf{x}_K) = \Phi \mathbf{X} \quad (3)$$

where $\mathbf{Y} \in \mathbb{R}^{m \times K}$ and $\mathbf{X} \in \mathbb{R}^{n \times K}$. Note that Equation (3) also rewrites $\mathbf{y}_k = \Phi \mathbf{x}_k$ for $1 \leq k \leq K$. The notion of support is extended to the matrix \mathbf{X} by defining $\text{supp}(\mathbf{X}) := \cup_{k \in [K]} \text{supp}(\mathbf{x}_k)$. Before introducing SOMP, we present some conventions.

Conventions: We consider the norms $\|\mathbf{x}\|_\infty := \max_{j \in [n]} |x_j|$ and $\|\mathbf{x}\|_p := (\sum_{j=1}^n |x_j|^p)^{1/p}$ where $1 \leq p < \infty$ and $\mathbf{x} \in \mathbb{R}^n$. In this work, any vector is a column vector. For $\mathcal{S} \subseteq [n]$, the vector $\mathbf{x}_{\mathcal{S}}$ is formed by the entries of \mathbf{x} whose indices belong to \mathcal{S} . In a likewise fashion, $\Phi_{\mathcal{S}}$ is defined as the matrix formed by the columns of Φ indexed within \mathcal{S} . Similarly, $\mathbf{X}^{\mathcal{S}}$ contains the rows of \mathbf{X} indexed by \mathcal{S} . The Moore-Penrose pseudoinverse, transpose, and conjugate transpose of any matrix Φ are denoted by Φ^+ , Φ^T , and Φ^* , respectively. The range of Φ is written $\mathcal{R}(\Phi)$. Also, the inner product of two vectors \mathbf{x} and \mathbf{y} is equal to $\langle \mathbf{x}, \mathbf{y} \rangle := \mathbf{x}^T \mathbf{y} = \mathbf{y}^T \mathbf{x}$. It is also worth defining the matrix norms $\|\mathbf{A}\|_{p \rightarrow q} := \sup_{\|\mathbf{z}\|_p=1} \|\mathbf{A}\mathbf{z}\|_q$. For $\mathbf{A} \in \mathbb{R}^{n \times K}$, we have $\|\mathbf{A}\|_{\infty \rightarrow \infty} = \max_{j \in [n]} \sum_{k=1}^K |A_{j,k}|$ as well as $\|\mathbf{A}\|_{2 \rightarrow 2} = \sqrt{\lambda_{\max}(\mathbf{A}^* \mathbf{A})}$ [16, Lemma A.5] where λ_{\max} denotes the maximal eigenvalue. Finally, the Frobenius norm of \mathbf{A} is denoted by $\|\mathbf{A}\|_F$.

1.2 Orthogonal Matching Pursuit algorithms

We present in this section the class of OMP algorithms for SMV and MMV models. In the event where the sparse signals \mathbf{x}_k to be recovered happen to share similar if not identical supports, it is interesting to perform a joint support recovery [18], *i.e.*, a single and common support $\hat{\mathcal{S}}$ is jointly estimated for all the K signals \mathbf{x}_k . SOMP [22], which is described in Algorithm 1, performs a joint support recovery. This algorithm iteratively picks atoms within Φ to simultaneously approximate the K measurement vectors \mathbf{y}_k . SOMP reduces to orthogonal matching pursuit (OMP) [10, 21] for $K = 1$.

At each iteration t , SOMP adds one atom to the estimated support (step 5). The criterion to determine which atom to include is to pick the atom maximizing $\|(\mathbf{R}^{(t)})^T \phi_j\|_1 =$

ALGORITHM 1:
Simultaneous orthogonal matching pursuit (SOMP)

Require: $\mathbf{Y} \in \mathbb{R}^{m \times K}$, $\Phi \in \mathbb{R}^{m \times n}$, $s \geq 1$

- 1: Initialization: $\mathbf{R}^{(0)} \leftarrow \mathbf{Y}$ and $\mathcal{S}_0 \leftarrow \emptyset$
- 2: $t \leftarrow 0$
- 3: **while** $t < s$ **do**
- 4: Determine the atom of Φ to be included in the support:
 $j_t \leftarrow \operatorname{argmax}_{j \in [n]} (\|(\mathbf{R}^{(t)})^T \phi_j\|_1)$
- 5: Update the support : $\mathcal{S}_{t+1} \leftarrow \mathcal{S}_t \cup \{j_t\}$
- 6: Projection of each measurement vector onto $\mathcal{R}(\Phi_{\mathcal{S}_{t+1}})$:
 $\mathbf{Y}^{(t+1)} \leftarrow \Phi_{\mathcal{S}_{t+1}} \Phi_{\mathcal{S}_{t+1}}^+ \mathbf{Y}$
- 7: Projection of each measurement vector onto $\mathcal{R}(\Phi_{\mathcal{S}_{t+1}})^\perp$:
 $\mathbf{R}^{(t+1)} \leftarrow \mathbf{Y} - \mathbf{Y}^{(t+1)}$
- 8: $t \leftarrow t + 1$
- 9: **end while**
- 10: **return** \mathcal{S}_s {Support at last step}

$\sum_{k=1}^K |\langle \mathbf{r}_k^{(t)}, \phi_j \rangle|$ (step 4) for the current residual matrix $\mathbf{R}^{(t)}$ where $\mathbf{r}_k^{(t)}$ denotes the k th column of $\mathbf{R}^{(t)}$. Note that the previous sum is a way to simultaneously account for all the measurement vectors \mathbf{y}_k and their corresponding residuals $\mathbf{r}_k^{(t)}$. The residual is then updated so that it is orthogonal to the subspace spanned by the atoms indexed by the current estimated support (steps 5 and 6.) The orthogonal projection matrix $\mathbf{P}^{(t)} := \Phi_{\mathcal{S}_t} \Phi_{\mathcal{S}_t}^+$ allows to perform the projection onto $\mathcal{R}(\Phi_{\mathcal{S}_{t+1}})$, *i.e.*, the space spanned by the columns of $\Phi_{\mathcal{S}_{t+1}}$. Using the ℓ_1 -norm for the decision criterion of SOMP is not the only possible choice. Generally, p -SOMP refers to the variant of SOMP for which the ℓ_p -norm intervenes [18]. Unless otherwise specified, we assume that SOMP uses the ℓ_1 -norm. The algorithm finishes when the size of the estimated support reaches s . If possible, s is usually chosen close to $|\mathcal{S}|$.

1.3 Contribution and its connection with the noisy case

In this paper, we provide a novel RIP-based lower bound of the quantity $\|\Phi_{\mathcal{S}}^T \mathbf{R}^{(t)}\|_{\infty \rightarrow \infty} = \max_{j \in \mathcal{S}} (\|(\mathbf{R}^{(t)})^T \phi_j\|_1)$, which is the maximum SOMP metric among the correct atoms without noise. In particular, we are interested in a lower bound expressed as $\|\Phi_{\mathcal{S}}^T \mathbf{R}^{(t)}\|_{\infty \rightarrow \infty} \geq \psi \tau_X$ where ψ only depends on \mathcal{S} and Φ while τ_X is determined by \mathbf{X} . This type of bound typically intervenes in noisy analyses [2, 8, 12] of OMP or SOMP, *i.e.*, for the signal model $\mathbf{Y} = \Phi \mathbf{X} + \mathbf{E}$ where $\mathbf{E} = (\mathbf{e}_1, \dots, \mathbf{e}_K)$ is the noise term. For example, [12, Theorem 3] shows that, if each noise vector obeys $\mathbf{e}_k \sim \mathcal{N}(\mathbf{0}, \sigma_k^2 \mathbf{I}_{m \times m})$, then the probability of SOMP identifying at least one incorrect support entry during s iterations is upper bounded by $\gamma(s, |\mathcal{S}|, n) \exp(-\Delta \mathbb{E}^2 / (8 \|\sigma\|_2^2))$ where $\gamma(s, |\mathcal{S}|, n)$ increases with s , $|\mathcal{S}|$, and n . $\Delta \mathbb{E}$ can be written as $\Delta \mathbb{E} = (1 - 1/\Gamma) \psi \tau_X - \sqrt{2/\pi} \|\sigma\|_1$ where the lower bound $\psi \tau_X$ intervenes. The quantity Γ , which is not studied in the present paper, connects the highest noiseless SOMP metrics for correct and incorrect atoms, *i.e.*, Γ lower bounds the ratio $\|\Phi_{\mathcal{S}}^T \mathbf{R}^{(t)}\|_{\infty \rightarrow \infty} / \|\Phi_{\mathcal{S}}^T \mathbf{R}^{(t)}\|_{\infty \rightarrow \infty}$ (see [11] for more details). It is worth pointing out again that, despite their intervening in the noisy analysis, the quantities Γ , ψ , and τ_X are all defined on the basis of noiseless signals.

1.4 Outline & related work overview

Section 2 explains in details the contribution. Section 3 states and comments an alternative bound in the literature. Finally, Section 4 compares our contribution against the alternative

lower bound. It is shown that, under several common sensing scenarios, our result outperforms its counterpart, even for $K = 1$, *i.e.*, for OMP. Our contribution can be used in several theoretical analyses [2, 8] of OMP and SOMP in the noisy case by replacing the older bound of Section 3 (Theorem 2) by the one obtained in this paper, *i.e.*, Theorem 1. For example, [8, Lemma 4.1] can be partially replaced by Theorem 1. The cases under which this replacement leads to less stringent conditions on whether an iteration is successful are hence discussed in Section 4. Other related algorithms include CoSaMP [20], Subspace pursuit [7], and orthogonal matching pursuit with replacement (OMPR) [19]. Since the decisions of these algorithms also rely on the highest inner products of a residual and the atoms, our methodology might provide relevant insights for them as well.

2 Contribution

Lemma 1 is an upper bound on $\|\mathbf{A}\|_{\infty \rightarrow \infty}$ depending on the distribution of the eigenvalues of \mathbf{A} . This lemma is needed in the proof of our main contribution, *i.e.*, Theorem 1.

Lemma 1. *Let $\alpha \in \mathbb{R}$ and $\mathbf{A} \in \mathbb{R}^{d \times d}$ be a normal matrix, *i.e.*, $\mathbf{A}\mathbf{A}^T = \mathbf{A}^T\mathbf{A}$. Let the vector $\boldsymbol{\theta}(\alpha)$ be composed of the elements $\theta_j(\alpha) = \lambda_j - \alpha$ where λ_j is the j th eigenvalue of \mathbf{A} . Then, $\|\mathbf{A}\|_{\infty \rightarrow \infty} \leq |\alpha| + \sqrt{d}\|\boldsymbol{\theta}(\alpha)\|_{\infty}$.*

Proof. The spectral theorem establishes that normal matrices are unitarily diagonalizable. We thus consider the eigenvalue decomposition $\mathbf{A} = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^*$ where \mathbf{Q} is unitary. Defining $\boldsymbol{\Delta}(\alpha) := \text{diag}(\boldsymbol{\theta}(\alpha))$, we have $\boldsymbol{\Lambda} = \alpha\mathbf{I} + \boldsymbol{\Delta}(\alpha)$. Thus, applying the triangle inequality yields

$$\|\mathbf{A}\|_{\infty \rightarrow \infty} \leq |\alpha| \underbrace{\|\mathbf{Q}\mathbf{Q}^*\|_{\infty \rightarrow \infty}}_{=1} + \|\mathbf{Q}\boldsymbol{\Delta}(\alpha)\mathbf{Q}^*\|_{\infty \rightarrow \infty}.$$

For $\mathbf{B} \in \mathbb{R}^{d \times d}$, the inequality $\|\mathbf{B}\|_{\infty \rightarrow \infty} \leq \sqrt{d}\|\mathbf{B}\|_{2 \rightarrow 2}$ [17] provides $\|\mathbf{A}\|_{\infty \rightarrow \infty} \leq |\alpha| + \sqrt{d}\|\mathbf{Q}\boldsymbol{\Delta}(\alpha)\mathbf{Q}^*\|_{2 \rightarrow 2} = |\alpha| + \sqrt{d}\|\boldsymbol{\theta}(\alpha)\|_{\infty}$. \square

Making use of the inequality derived in Lemma 1, Theorem 1 states our novel lower bound on $\|\boldsymbol{\Phi}_S^T \mathbf{R}^{(t)}\|_{\infty \rightarrow \infty}$.

Theorem 1. *Let us assume that SOMP has picked only correct atoms before iteration t , *i.e.*, $\mathcal{S}_t \subset \mathcal{S}$. We denote $\mathcal{J}_t = \mathcal{S} \setminus \mathcal{S}_t$ the set that contains the indices of the correct atoms yet to be selected at iteration t . If $\boldsymbol{\Phi}$ satisfies the RIP with $|\mathcal{S}|$ -th RIC $\delta_{|\mathcal{S}|} < 1$, then*

$$\|\boldsymbol{\Phi}_S^T \mathbf{R}^{(t)}\|_{\infty \rightarrow \infty} \geq \frac{(1 - \delta_{|\mathcal{S}|})(1 + \delta_{|\mathcal{S}|})}{1 + \sqrt{|\mathcal{S}| - t} \delta_{|\mathcal{S}|}} \|\mathbf{X}^{\mathcal{J}_t}\|_{\infty \rightarrow \infty}. \quad (4)$$

Proof. We have $\boldsymbol{\Phi}_S^T \mathbf{R}^{(t)} = \boldsymbol{\Phi}_S^T (\mathbf{I} - \mathbf{P}^{(t)}) \boldsymbol{\Phi}_{\mathcal{S}} \mathbf{X}^{\mathcal{S}} = \boldsymbol{\Phi}_S^T (\mathbf{I} - \mathbf{P}^{(t)}) \boldsymbol{\Phi}_{\mathcal{J}_t} \mathbf{X}^{\mathcal{J}_t}$ because $(\mathbf{I} - \mathbf{P}^{(t)}) \boldsymbol{\Phi}_{\mathcal{S}} \mathbf{X}^{\mathcal{S}} = (\mathbf{I} - \mathbf{P}^{(t)}) (\boldsymbol{\Phi}_{\mathcal{S}_t} \mathbf{X}^{\mathcal{S}_t} + \boldsymbol{\Phi}_{\mathcal{J}_t} \mathbf{X}^{\mathcal{J}_t}) = (\mathbf{I} - \mathbf{P}^{(t)}) \boldsymbol{\Phi}_{\mathcal{J}_t} \mathbf{X}^{\mathcal{J}_t}$. For $j \in \mathcal{S}_t$, $\langle \phi_j, (\mathbf{I} - \mathbf{P}^{(t)}) \mathbf{z} \rangle = 0$ for every vector \mathbf{z} . Thus,

$$\begin{aligned} \|\boldsymbol{\Phi}_S^T \mathbf{R}^{(t)}\|_{\infty \rightarrow \infty} &= \|\boldsymbol{\Phi}_S^T (\mathbf{I} - \mathbf{P}^{(t)}) \boldsymbol{\Phi}_{\mathcal{J}_t} \mathbf{X}^{\mathcal{J}_t}\|_{\infty \rightarrow \infty} \\ &= \max_{j \in \mathcal{S}} \sum_{k=1}^K |\langle \phi_j, (\mathbf{I} - \mathbf{P}^{(t)}) (\boldsymbol{\Phi}_{\mathcal{J}_t} \mathbf{X}^{\mathcal{J}_t})_k \rangle| \\ &= \max_{j \in \mathcal{J}_t} \sum_{k=1}^K |\langle \phi_j, (\mathbf{I} - \mathbf{P}^{(t)}) (\boldsymbol{\Phi}_{\mathcal{J}_t} \mathbf{X}^{\mathcal{J}_t})_k \rangle| \\ &= \|\boldsymbol{\Phi}_{\mathcal{J}_t}^T (\mathbf{I} - \mathbf{P}^{(t)}) \boldsymbol{\Phi}_{\mathcal{J}_t} \mathbf{X}^{\mathcal{J}_t}\|_{\infty \rightarrow \infty}. \end{aligned}$$

We follow the steps of [18, Proof of Theorem 10] and use the inequality $\|\mathbf{A}\mathbf{B}\|_{\infty \rightarrow \infty} \leq \|\mathbf{A}\|_{\infty \rightarrow \infty} \|\mathbf{B}\|_{\infty \rightarrow \infty}$. If $\mathbf{B} = \mathbf{C}\mathbf{D}$ and \mathbf{C} is invertible, then, with $\mathbf{A} = \mathbf{C}^{-1}$, the inequality above implies $\|\mathbf{C}^{-1}\mathbf{C}\mathbf{D}\|_{\infty \rightarrow \infty} = \|\mathbf{D}\|_{\infty \rightarrow \infty} \leq \|\mathbf{C}^{-1}\|_{\infty \rightarrow \infty} \|\mathbf{C}\mathbf{D}\|_{\infty \rightarrow \infty}$. Replacing \mathbf{D} with $\mathbf{X}^{\mathcal{J}_t}$ and \mathbf{C} with $\Phi_{\mathcal{J}_t}^T(\mathbf{I} - \mathbf{P}^{(t)})\Phi_{\mathcal{J}_t}$ yields

$$\|\Phi_{\mathcal{S}}^T \mathbf{R}^{(t)}\|_{\infty \rightarrow \infty} \geq \frac{\|\mathbf{X}^{\mathcal{J}_t}\|_{\infty \rightarrow \infty}}{\|(\Phi_{\mathcal{J}_t}^T(\mathbf{I} - \mathbf{P}^{(t)})\Phi_{\mathcal{J}_t})^{-1}\|_{\infty \rightarrow \infty}}$$

It can be shown [2, Lemma 5] that, under the condition $\mathcal{S}_t \subseteq \mathcal{S}$, we have $\lambda_{\min}(\Phi_{\mathcal{J}_t}^T(\mathbf{I} - \mathbf{P}^{(t)})\Phi_{\mathcal{J}_t}) \geq \lambda_{\min}(\Phi_{\mathcal{S}}^T\Phi_{\mathcal{S}}) \geq 1 - \delta_{|\mathcal{S}|} > 0$ (see [7, Remark 1] regarding the penultimate inequality). The absence of zero eigenvalues thus shows that the matrix $\Phi_{\mathcal{J}_t}^T(\mathbf{I} - \mathbf{P}^{(t)})\Phi_{\mathcal{J}_t}$ is full rank and invertible. Note that it can also be shown [2, Lemma 5] that $\lambda_{\max}(\Phi_{\mathcal{J}_t}^T(\mathbf{I} - \mathbf{P}^{(t)})\Phi_{\mathcal{J}_t}) \leq \lambda_{\max}(\Phi_{\mathcal{S}}^T\Phi_{\mathcal{S}}) \leq 1 + \delta_{|\mathcal{S}|}$ (see [7, Remark 1] for the last inequality). As a result, the eigenvalues of $(\Phi_{\mathcal{J}_t}^T(\mathbf{I} - \mathbf{P}^{(t)})\Phi_{\mathcal{J}_t})^{-1}$ belong to $[1/(1 + \delta_{|\mathcal{S}|}); 1/(1 - \delta_{|\mathcal{S}|})]$. Due to the idempotency and symmetry of the orthogonal projectors, $\Phi_{\mathcal{J}_t}^T(\mathbf{I} - \mathbf{P}^{(t)})\Phi_{\mathcal{J}_t} = \Phi_{\mathcal{J}_t}^T(\mathbf{I} - \mathbf{P}^{(t)})^T(\mathbf{I} - \mathbf{P}^{(t)})\Phi_{\mathcal{J}_t}$, which implies that the matrix of interest and its inverse are symmetric and consequently normal.

Instead of using the coherence of Φ as in [18], we consider the eigenvalues of $(\Phi_{\mathcal{J}_t}^T(\mathbf{I} - \mathbf{P}^{(t)})\Phi_{\mathcal{J}_t})^{-1}$. Since this last matrix is normal, Lemma 1 can be used with

$$\alpha = \frac{1}{2} \left(\frac{1}{1 + \delta_{|\mathcal{S}|}} + \frac{1}{1 - \delta_{|\mathcal{S}|}} \right) = \frac{1}{(1 + \delta_{|\mathcal{S}|})(1 - \delta_{|\mathcal{S}|})},$$

which is the arithmetic mean of the lowest and highest possible eigenvalues of $(\Phi_{\mathcal{J}_t}^T(\mathbf{I} - \mathbf{P}^{(t)})\Phi_{\mathcal{J}_t})^{-1}$ given the RIP with the RIC $\delta_{|\mathcal{S}|}$. For such a choice,

$$\|\boldsymbol{\theta}(\alpha)\|_{\infty} \leq \frac{1}{1 - \delta_{|\mathcal{S}|}} - \alpha = \frac{\delta_{|\mathcal{S}|}}{(1 + \delta_{|\mathcal{S}|})(1 - \delta_{|\mathcal{S}|})}.$$

Thus, Lemma 1 yields

$$\|(\Phi_{\mathcal{J}_t}^T(\mathbf{I} - \mathbf{P}^{(t)})\Phi_{\mathcal{J}_t})^{-1}\|_{\infty \rightarrow \infty} \leq \frac{1 + \sqrt{|\mathcal{J}_t|}\delta_{|\mathcal{S}|}}{(1 + \delta_{|\mathcal{S}|})(1 - \delta_{|\mathcal{S}|})},$$

which concludes the proof. \square

Theorem 1 essentially states that the maximal correlation obtained among the correct atoms, *i.e.*, $\|\Phi_{\mathcal{S}}^T \mathbf{R}^{(t)}\|_{\infty \rightarrow \infty}$, is lower bounded by a quantity proportional to $\|\mathbf{X}^{\mathcal{J}_t}\|_{\infty \rightarrow \infty} = \max_{j \in \mathcal{J}_t} \|\mathbf{X}^{\{j\}}\|_1 = \max_{j \in \mathcal{J}_t} \sum_{k=1}^K |X_{j,k}|$. As already stated in Section 1.3, lower bounds on $\|\Phi_{\mathcal{S}}^T \mathbf{R}^{(t)}\|_{\infty \rightarrow \infty}$ in the noiseless case play a role when determining the performance of SOMP when additive noise is included in the signal model. The properties (including the sharpness) of Theorem 1 will be discussed in Section 4

3 Related work

Let us now compare our lower bound on $\|\Phi_{\mathcal{S}}^T \mathbf{R}^{(t)}\|_{\infty \rightarrow \infty}$, *i.e.*, Theorem 1, to another important one in the literature, *i.e.*, Theorem 2. To the best of the authors' knowledge, Theorem 2 was first obtained in [23, Section 3.1] for 2-SOMP. In the SMV case, the inequality $\|\Phi_{\mathcal{S}}^T \mathbf{r}^{(t)}\|_{\infty} \geq \frac{\lambda_{\min}(\Phi_{\mathcal{S}}^T\Phi_{\mathcal{S}})}{\sqrt{|\mathcal{S}| - t}} \|\mathbf{x}_{\mathcal{J}_t}\|_2$ was first obtained in [2, Section V] for OMP, which immediately yields Theorem 2 for $K = 1$ when using the inequality $\lambda_{\min}(\Phi_{\mathcal{S}}^T\Phi_{\mathcal{S}}) \geq 1 - \delta_{|\mathcal{S}|}$ [7, Remark 1].

Theorem 2. Let us assume that SOMP has picked only correct atoms before iteration t , i.e., $\mathcal{S}_t \subset \mathcal{S}$, with $\mathcal{J}_t = \mathcal{S} \setminus \mathcal{S}_t$ containing the indices of the correct atoms yet to be selected at iteration t . If Φ satisfies the RIP with $|\mathcal{S}|$ -th RIC $\delta_{|\mathcal{S}|} < 1$, then

$$\|\Phi_{\mathcal{S}}^T \mathbf{R}^{(t)}\|_{\infty \rightarrow \infty} \geq (1 - \delta_{|\mathcal{S}|}) \frac{1}{\sqrt{|\mathcal{S}| - t}} \|\mathbf{X}^{\mathcal{J}_t}\|_{\text{F}}. \quad (5)$$

Proof. Rearranging the results in [23, Section 3.1] shows that Theorem 2 is true if $\max_{j \in \mathcal{S}} \|(\mathbf{R}^{(t)})^T \phi_j\|_1$ is replaced by $\max_{j \in \mathcal{S}} \|(\mathbf{R}^{(t)})^T \phi_j\|_2$, i.e., if 2-SOMP is used instead of 1-SOMP. Since $\|\mathbf{x}\|_1 \geq \|\mathbf{x}\|_2$ for all \mathbf{x} , $\|\Phi_{\mathcal{S}}^T \mathbf{R}^{(t)}\|_{\infty \rightarrow \infty} = \max_{j \in \mathcal{S}} \|(\mathbf{R}^{(t)})^T \phi_j\|_1 \geq \max_{j \in \mathcal{S}} \|(\mathbf{R}^{(t)})^T \phi_j\|_2 \geq (1 - \delta_{|\mathcal{S}|}) \|\mathbf{X}^{\mathcal{J}_t}\|_{\text{F}} / \sqrt{|\mathcal{S}| - t}$, which concludes the proof. \square

The quantity $\|\mathbf{X}^{\mathcal{J}_t}\|_{\text{F}}^2$ rewrites $\sum_{j \in \mathcal{J}_t} \|\mathbf{X}^{\{j\}}\|_2^2$, which is the sum of the squared ℓ_2 -norms of each row of \mathbf{X} indexed by \mathcal{J}_t . Since each row of \mathbf{X} can be interpreted as the coefficient vector associated with one particular atom, $\|\mathbf{X}^{\mathcal{J}_t}\|_{\text{F}}^2 / (|\mathcal{S}| - t)$ is the average energy of the coefficients associated with the atoms indexed by \mathcal{J}_t .

4 Comparison with related works & Discussions

The rest of this section is dedicated to the comparison of Theorem 1, i.e., our contribution, with Theorem 2. To determine which bound is the better at iteration t , we introduce the following quantity

$$r(|\mathcal{S}|, \mathcal{J}_t) := \frac{\sqrt{|\mathcal{J}_t|}(1 + \delta_{|\mathcal{S}|})}{1 + \sqrt{|\mathcal{J}_t|}\delta_{|\mathcal{S}|}} \frac{\|\mathbf{X}^{\mathcal{J}_t}\|_{\infty \rightarrow \infty}}{\|\mathbf{X}^{\mathcal{J}_t}\|_{\text{F}}}, \quad (6)$$

which is the ratio of the lower bound of Theorem 1 to that of Theorem 2. Our contribution thereby improves the analysis of SOMP when $r(|\mathcal{S}|, \mathcal{J}_t) > 1$. For $\mathbf{B} \in \mathbb{R}^{|\mathcal{J}_t| \times K}$, we have $\|\mathbf{B}\|_{\infty \rightarrow \infty} \leq \sqrt{K} \|\mathbf{B}\|_{2 \rightarrow 2} \leq \sqrt{K} \|\mathbf{B}\|_{\text{F}}$ [17]. For $\mathbf{b}_j := \|\mathbf{B}^{\{j\}}\|_2$ (where $\mathbf{b} \in \mathbb{R}^{|\mathcal{J}_t|}$), we obtain $\|\mathbf{B}\|_{\infty \rightarrow \infty} = \max_{j \in [\mathcal{J}_t]} \|\mathbf{B}^{\{j\}}\|_1 \geq \max_{j \in [\mathcal{J}_t]} \|\mathbf{B}^{\{j\}}\|_2 = \|\mathbf{b}\|_{\infty} \geq (1/\sqrt{|\mathcal{J}_t|}) \|\mathbf{b}\|_2 = (1/\sqrt{|\mathcal{J}_t|}) \|\mathbf{B}\|_{\text{F}}$. As a result,

$$\begin{aligned} \frac{1 + \delta_{|\mathcal{S}|}}{1 + \sqrt{|\mathcal{J}_t|}\delta_{|\mathcal{S}|}} &\leq r(|\mathcal{S}|, \mathcal{J}_t) \leq \frac{(1 + \delta_{|\mathcal{S}|})\sqrt{|\mathcal{J}_t|}\sqrt{K}}{1 + \sqrt{|\mathcal{J}_t|}\delta_{|\mathcal{S}|}} \\ &\leq \sqrt{K} \frac{1 + \delta_{|\mathcal{S}|}}{\delta_{|\mathcal{S}|}}. \end{aligned}$$

The proposed analysis of $r(|\mathcal{S}|, \mathcal{J}_t)$ is realized for four different reconstruction scenarios that are discussed hereafter.

4.1 Case 1: A single dominant row within \mathbf{X}

We assume that $\|\mathbf{X}^{\mathcal{J}_t}\|_{\infty \rightarrow \infty} \simeq \|\mathbf{X}^{\{j_d\}}\|_1$ and $\|\mathbf{X}^{\mathcal{J}_t}\|_{\text{F}} \simeq \|\mathbf{X}^{\{j_d\}}\|_2$ for some $j_d \in \mathcal{S}$. This situation occurs whenever the entries of the j_d th row of \mathbf{X} have magnitudes overwhelmingly higher than those of all the other rows combined. Then,

$$r(|\mathcal{S}|, \mathcal{J}_t) \simeq \frac{\sqrt{|\mathcal{J}_t|}(1 + \delta_{|\mathcal{S}|})}{1 + \sqrt{|\mathcal{J}_t|}\delta_{|\mathcal{S}|}} \frac{\|\mathbf{X}^{\{j_d\}}\|_1}{\|\mathbf{X}^{\{j_d\}}\|_2} \geq \frac{\sqrt{|\mathcal{J}_t|}(1 + \delta_{|\mathcal{S}|})}{1 + \sqrt{|\mathcal{J}_t|}\delta_{|\mathcal{S}|}} \quad (7)$$

where $1 \leq \|\mathbf{x}\|_1 / \|\mathbf{x}\|_2 \leq \sqrt{K}$ for all $\mathbf{x} \in \mathbb{R}^K$. Theorem 1 always outperforms Theorem 2 in this case since the RHS of Equation (7) is higher than 1. Interestingly, it remains true

in the SMV setting, thereby making our contribution superior to the state-of-the-art result of [2] for OMP. If $\mathbf{X}^{\{j_d\}}$ is 1-sparse, then $r(|\mathcal{S}|, \mathcal{J}_t) = \sqrt{|\mathcal{J}_t|}(1 + \delta_{|\mathcal{S}|})/(1 + \sqrt{|\mathcal{J}_t|} \delta_{|\mathcal{S}|})$ while $r(|\mathcal{S}|, \mathcal{J}_t) = \sqrt{K}\sqrt{|\mathcal{J}_t|}(1 + \delta_{|\mathcal{S}|})/(1 + \sqrt{|\mathcal{J}_t|} \delta_{|\mathcal{S}|})$ whenever the entries of $\mathbf{X}^{\{j_d\}}$ have identical absolute values. These first observations suggest that the improvements resulting from using Theorem 1 instead of Theorem 2 increase with K .

4.2 Case 2: Identical magnitudes

We assume that $|X_{j,k}| \simeq \mu_X > 0$ for each $(j, k) \in \mathcal{J}_t \times [K]$. Thus, we have $\|\mathbf{X}^{\mathcal{J}_t}\|_{\infty \rightarrow \infty} \simeq K\mu_X$ and $\|\mathbf{X}^{\mathcal{J}_t}\|_F \simeq \sqrt{K}\sqrt{|\mathcal{J}_t|}\mu_X$. As a result, we obtain

$$r(|\mathcal{S}|, \mathcal{J}_t) \simeq \frac{\sqrt{K}(1 + \delta_{|\mathcal{S}|})}{1 + \sqrt{|\mathcal{J}_t|} \delta_{|\mathcal{S}|}}. \quad (8)$$

As depicted in Figure 1, the situation might be favorable to both lower bounds depending on the value of K , $|\mathcal{J}_t|$, and $\delta_{|\mathcal{S}|}$. In this case, Theorem 1 is always worse than its counterpart if $K = 1$, $|\mathcal{J}_t| > 1$, and $\delta_{|\mathcal{S}|} > 0$. As a general rule, our contribution tends to get better than Theorem 2 whenever K increases, the RIC $\delta_{|\mathcal{S}|}$ approaches 0, or the number of correct atoms yet to be recovered, *i.e.*, $|\mathcal{J}_t|$, tends to 1.

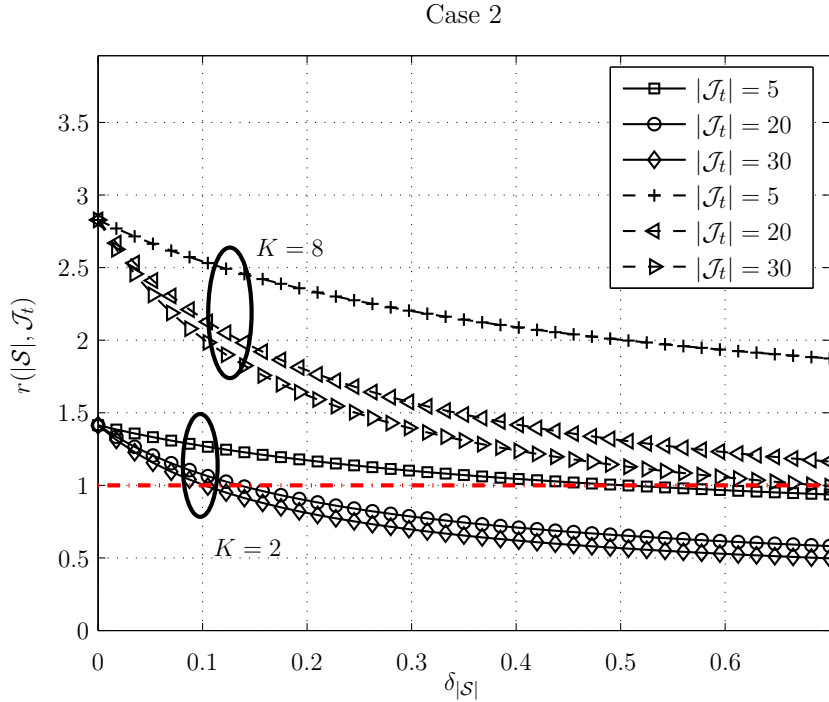


Figure 1: Analysis of $r(|\mathcal{S}|, \mathcal{J}_t)$ in Case 2 for various 2-tuples $(|\mathcal{J}_t|, \delta_{\mathcal{S}})$ – The dot-dash horizontal red line determines when both theorems provide equivalent bounds, *i.e.*, $r(|\mathcal{S}|, \mathcal{J}_t) = 1$.

4.3 Case 3: Last iteration

We assume that only one correct atom has yet to be picked, *i.e.*, $\mathcal{J}_t = \{j_f\}$. Thus,

$$r(|\mathcal{S}|, \mathcal{J}_t) = \frac{\|\mathbf{X}^{\{j_f\}}\|_1}{\|\mathbf{X}^{\{j_f\}}\|_2} \geq 1. \quad (9)$$

If the row $\mathbf{X}^{\{j_f\}}$ is 1-sparse, then $r(|\mathcal{S}|, \mathcal{J}_t) = 1$ while $r(|\mathcal{S}|, \mathcal{J}_t) = \sqrt{K}$ whenever its entries have identical absolute values. Both theorems deliver the same performance within the SMV framework but differ as soon as $K > 1$.

4.4 Case 4: “Perfect” measurement matrix

Let us assume that $\delta_{|\mathcal{S}|} = 0$. Hence,

$$r(|\mathcal{S}|, \mathcal{J}_t) = \sqrt{|\mathcal{J}_t|} \frac{\|\mathbf{X}^{\mathcal{J}_t}\|_{\infty \rightarrow \infty}}{\|\mathbf{X}^{\mathcal{J}_t}\|_F} \geq 1. \quad (10)$$

Our contribution is at least equivalent to the state-of-the-art bound in this case. As stated in Remark 1, Theorem 1 is sharp for $\delta_{|\mathcal{S}|} = 0$, *i.e.*, $\|\Phi_{\mathcal{S}}^T \mathbf{R}^{(t)}\|_{\infty \rightarrow \infty} = \|\mathbf{X}^{\mathcal{J}_t}\|_{\infty \rightarrow \infty}$.

Remark 1. $\delta_{|\mathcal{S}|} = 0 \Rightarrow \|\Phi_{\mathcal{S}}^T \mathbf{R}^{(t)}\|_{\infty \rightarrow \infty} = \|\mathbf{X}^{\mathcal{J}_t}\|_{\infty \rightarrow \infty}$.

Proof. For each $j_1, j_2 \in \mathcal{S}$ such that $j_1 \neq j_2$, we have $|\langle \phi_{j_1}, \phi_{j_2} \rangle| \leq \delta_2 \|\phi_{j_1}\|_2 \|\phi_{j_2}\|_2$ (see [4, Lemma 2.1]) where $\delta_2 \leq \delta_{|\mathcal{S}|} = 0$ because the RIC is monotonically increasing, *i.e.*, $\delta_s \leq \delta_{s+1}$. Hence, all the atoms comprised within $\Phi_{\mathcal{S}}$ are orthogonal to each other. Similarly to the proof of Theorem 1, we have $\Phi_{\mathcal{S}}^T \mathbf{R}^{(t)} = \Phi_{\mathcal{S}}^T (\mathbf{I} - \mathbf{P}^{(t)}) \Phi_{\mathcal{S}} \mathbf{X}^{\mathcal{S}} = \Phi_{\mathcal{S}}^T \Phi_{\mathcal{J}_t} \mathbf{X}^{\mathcal{J}_t}$ and $\|\Phi_{\mathcal{S}}^T \mathbf{R}^{(t)}\|_{\infty \rightarrow \infty} = \|\Phi_{\mathcal{J}_t}^T \Phi_{\mathcal{J}_t} \mathbf{X}^{\mathcal{J}_t}\|_{\infty \rightarrow \infty}$. The vanishing of the orthogonal projection matrix stems from the orthogonality of the atoms indexed by \mathcal{S} . The matrix $(\mathbf{I} - \mathbf{P}^{(t)})$ indeed projects onto $\mathcal{R}(\Phi_{\mathcal{S}_t})^\perp$ and $\mathcal{R}(\Phi_{\mathcal{J}_t}) \subset \mathcal{R}(\Phi_{\mathcal{S}_t})^\perp$ since $\mathcal{R}(\Phi_{\mathcal{S}_t}) \perp \mathcal{R}(\Phi_{\mathcal{J}_t})$. As $\delta_{|\mathcal{S}|} = 0$, we have [16, Equation 6.2] $\|\Phi_{\mathcal{J}_t}^T \Phi_{\mathcal{J}_t} - \mathbf{I}\|_{2 \rightarrow 2} = 0$ so that $\Phi_{\mathcal{J}_t}^T \Phi_{\mathcal{J}_t} = \mathbf{I}$. \square

Acknowledgments

The authors would like to thank the Belgian “Fonds de la recherche scientifique” for having funded this research.

References

- [1] D. Baron, M. F. Duarte, M. B. Wakin, S. Sarvotham, and R. G. Baraniuk. Distributed compressed sensing. *Rice University Technical Report*, (TREE-0612), Nov. 2006.
- [2] T. T. Cai and L. Wang. Orthogonal matching pursuit for sparse signal recovery with noise. *Information Theory, IEEE Transactions on*, 57(7):4680–4688, Jul. 2011.
- [3] T. T. Cai, L. Wang, and G. Xu. Shifting inequality and recovery of sparse signals. *IEEE Trans. Signal Process.*, 58(3):1300–1308, Mar. 2010.
- [4] E. J. Candès. The restricted isometry property and its implications for compressed sensing. *Comptes Rendus Mathématique*, 346(9):589–592, May 2008.
- [5] E. J. Candes, J. K. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on pure and applied mathematics*, 59(8):1207–1223, Aug. 2006.
- [6] E. J. Candes and T. Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. Inf. Theory*, 52(12):5406–5425, Dec. 2006.
- [7] W. Dai and O. Milenkovic. Subspace pursuit for compressive sensing signal reconstruction. *IEEE Trans. Inf. Theory*, 55(5):2230–2249, May 2009.
- [8] W. Dan and R. Wang. Robustness of orthogonal matching pursuit under restricted isometry property. *Science China Mathematics*, 57(3):627–634, Mar. 2014.
- [9] M. Davenport, M. B. Wakin, et al. Analysis of orthogonal matching pursuit using the restricted isometry property. *IEEE Trans. Inf. Theory*, 56(9):4395–4401, Sep. 2010.
- [10] G. Davis, S. Mallat, and M. Avellaneda. Adaptive greedy approximations. *Constructive approximation*, 13(1):57–98, Mar. 1997.
- [11] J.-F. Determe, J. Louveaux, L. Jacques, and F. Horlin. On the exact recovery condition of simultaneous orthogonal matching pursuit. *Signal Processing Letters, IEEE*, 23(1):164–168, Jan. 2016.
- [12] J.-F. Determe, J. Louveaux, L. Jacques, and F. Horlin. On the noise robustness of simultaneous orthogonal matching pursuit. *arXiv preprint arXiv:1602.03400*, 2016.
- [13] J. Ding, L. Chen, and Y. Gu. Robustness of orthogonal matching pursuit for multiple measurement vectors in noisy scenario. In *Proc. 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3813–3816, Kyoto International Conference Center (KICC), Kyoto, Japan, Mar. 2012. IEEE.
- [14] D. L. Donoho. Compressed sensing. *IEEE Trans. Inf. Theory*, 52(4):1289–1306, Apr. 2006.
- [15] Y. C. Eldar and M. Mishali. Robust recovery of signals from a structured union of subspaces. *IEEE Trans. Inf. Theory*, 55(11):5302–5316, Nov. 2009.
- [16] S. Foucart and H. Rauhut. *A mathematical introduction to compressive sensing*. Springer, Aug. 2013.
- [17] G. H. Golub and C. F. Van Loan. *Matrix computations*, volume 3. JHU Press, Dec. 2012.

- [18] R. Gribonval, H. Rauhut, K. Schnass, and P. Vandergheynst. Atoms of all channels, unite! Average case analysis of multi-channel sparse recovery using greedy algorithms. *Journal of Fourier analysis and Applications*, 14(5-6):655–687, Dec. 2008.
- [19] P. Jain, A. Tewari, and I. S. Dhillon. Orthogonal matching pursuit with replacement. In *Advances in Neural Information Processing Systems*, pages 1215–1223, 2011.
- [20] D. Needell and J. A. Tropp. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Applied and Computational Harmonic Analysis*, 26(3):301–321, May 2009.
- [21] Y. C. Pati, R. Rezaeiifar, and P. S. Krishnaprasad. Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition. In *Proc. 27th Annu. Asilomar Conf. Signals, Systems, and Computers*, volume 1, pages 40–44, Pacific Grove, CA, Nov. 1993. IEEE.
- [22] J. A. Tropp, A. C. Gilbert, and M. J. Strauss. Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit. *Signal Processing*, 86(3):572–588, Mar. 2006.
- [23] Y. Wang, T. Fu, M. Gao, and S. Ding. Performance of orthogonal matching pursuit for multiple measurement vectors with noise. In *Proc. 2013 IEEE China Summit International Conference on Signal and Information Processing (ChinaSIP)*, pages 67–71, China National Convention Center, Beijing, China, Jul. 2013. IEEE.